



Силабус освітнього компонента

Програма навчальної дисципліни

Програмне забезпечення інженерії даних і знань



Шифр та назва спеціальності
122 – Комп'ютерні науки

Інститут
ННІ Комп'ютерних наук та інформаційних
технологій

Освітня програма
Комп'ютерні науки

Кафедра
Системного аналізу та інформаційно-
аналітичних технологій (322)

Рівень освіти
Магістр

Тип дисципліни
Спеціальна (фахова), вибіркова

Семестр
2

Мова викладання
Українська

Викладачі, розробники



Коваленко Сергій Володимирович

Serhii.Kovalenko@khp.edu.ua

Кандидат технічних наук, доцент, професор кафедри системного аналізу та інформаційно-аналітичних технологій НТУ «ХПІ»

Досвід роботи – 27 років. Автор понад 100 наукових та навчально-методичних праць. Провідний лектор з дисциплін: «Проектування баз даних», «Обробка текстової та графічної інформації», «Бази даних», «Комп'ютерна обробка зображень», «Обробка зображень та мультимедіа», «Основи комп'ютерної графіки», «Обробка даних засобами Python».

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Дисципліна забезпечує засвоєння знань та навичок, які необхідні для використання сучасних аналітичних інструментів аналізу великих даних, опанування методів, моделей та алгоритмів аналізу даних засобами інтерактивних програмних платформ,

Мета та цілі дисципліни

Метою викладання дисципліни є формування у студентів загальних та професійних компетентностей з теорії та практики аналізу даних – від первинного збору даних до вибору оптимального рішення та оцінювання його якості, а також оволодіння методами, технологіями та засобами консолідації інформації та представлення знань.

Формат занять

Лекції, лабораторні роботи, самостійна робота, консультації. Підсумковий контроль – залік.

Компетентності

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.
ЗК02. Здатність застосовувати знання у практичних ситуаціях.
ЗК05. Здатність вчитися й оволодівати сучасними знаннями.

СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК04. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проектних рішень.

СК06. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.

Результати навчання

РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

РН8. Розробляти математичні моделі та методи аналізу даних (включно з великим).

РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

Обсяг дисципліни

Загальний обсяг дисципліни 120 год. (4 кредити ECTS): лекції – 32 год., лабораторні роботи – 16 год., самостійна робота – 72 год.

Передумови вивчення дисципліни (пререквізити)

Для успішного вивчення дисципліни необхідно мати знання та практичні навички з дисциплін "Інтелектуальний аналіз даних", "Сучасні методи математичного та комп'ютерного моделювання".

Особливості дисципліни, методи та технології навчання

Лекції проводяться з використанням мультимедійних технологій.

Навчальні матеріали доступні студентам через OneDrive кафедри.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Вступ до аналізу даних

Основи синтаксису Python. Інструменти Python для аналізу даних, дослідження та візуалізації функцій.

Тема 2. Використання спеціалізованих бібліотек Python

Інструменти Python, що реалізують операції лінійної алгебри. Вектори, матриці та операції над ними в бібліотеці NumPy. Бібліотеки NumPy, Matplotlib, SciPy, Pandas. DataFrame – інструменти бібліотеки Pandas, індексація та селекція. Вирішення оптимізаційних задач в SciPy.

Тема 3. Оптимізація функцій засобами Python

Інструменти Python для оптимізації гладких функцій. Інструменти Python для оптимізації негладких функцій. Метод імітації віджигу. Генетичні алгоритми та диференційна еволюція.

Тема 4. Інструменти Python для роботи з випадковими величинами

Генерація випадкових величин із різноманітними законами розподілу. Побудова графіків щільностей розподілу ймовірності та функцій розподілу ймовірності. Підрахунок ймовірності подій, які пов'язані із випадковими величинами із різноманітними законами розподілу.

Тема 5. Інструменти Python, які реалізують методи математичної статистики

Оцінка закону розподілу випадкової величини за наявною виборкою. Закони розподілу випадкових величин, параметри та оцінки. Центральна гранична теорема. Довірчі інтервали.

Тема 6. Інструменти Python для обчислення довірчих інтервалів за допомогою квантілів

Розподіли випадкових величин, які є похідними від нормального. Довірчі інтервали для середнього, для долі, для двох долей. Довірчі інтервали на основі бутстрепа.

Тема 7. Інструменти Python, що реалізують методи перевірки статистичних гіпотез

Помилки першого та другого роду, рівень значущості, статистична та практична значущість. Біноміальний критерій для долі. Критерій узгодженості Пірсона (χ^2 - квадрат). зв'язок між перевіркою гіпотез та довірчими інтервалами.

Тема 8. Інструменти Python, що реалізують АБ-тестування

Використання АБ-тестування, метрики, дизайн експерименту, стійкість, розмір вибірки.

Тема 9. Інструменти Python, що реалізують пошук закономірностей в даних

Кореляції Пірсона, Спірмена, Метью. Значимість кореляції. Булшит та консервативність. Кореляція та причинно-наслідковий зв'язок. Кореляційний аналіз.

Тема 10. Множинна перевірка гіпотез за допомогою інструментів Python

Постановка задачі. Групова ймовірність помилки першого роду (family-wise error rate, FWER). FWER поправка Боффероні. FWER Метод Холма. Метод Бенджаміні-Хохберга.

Тема 11. Приклади задач аналізу даних

Приклади вирішення задач аналізу даних за допомогою інструментів Python.

Теми практичних занять

Практичні заняття не передбачені

Теми лабораторних робіт

Тема 1.

Встановлення інструментів Anaconda Python на комп'ютер користувача. Створення документів іpython notebook. Практична робота в середовищі Anaconda Python.

Тема 2.

Використання засобів Python для аналізу схожості текстів та апроксимації функцій..

Тема 3.

Застосування засобів Python для вирішення задач глобальної оптимізації та оптимізація негладкої функції.

Тема 4.

Використання інструментів Python для генерації випадкових величин та підрахунку ймовірностей.

Тема 5.

Перевірка центральної граничної теореми за допомогою інструментів Python.

Тема 6.

Застосування бібліотек Python для вирішення практичних задач побудови довірчих інтервалів.

Тема 7.

Вирішення задач щодо перевірки статистичних гіпотез за допомогою інструментів Python.

Тема 8.

Аналіз результатів АБ-тесту у середовищі Anaconda Python.

Самостійна робота

Самостійна робота передбачає виконання таких тем та завдань:

1. Опрацювання лекційного матеріалу.
2. Виконання завдань, які пропонуються викладачем в якості домашніх завдань.
3. Підготовка до контрольних робіт.
4. Студентам рекомендуються додаткові матеріали для самостійного вивчення та аналізу.

Література та навчальні матеріали

Основна література

1. Florent Buisson. (2021). Behavioral Data Analysis with R and Python. USA: O'Reilly, 336 p. [Електронний ресурс] URL : <https://dokumen.pub/qdownload/behavioral-data-analysis-with-r-and-python-customer-driven-data-for-real-business-results-9781492061373-1492061379.html>
2. Jake VanderPlas. (2023). Python Data Science Handbook. 2-nd Edition USA: O'Reilly Media, 564 p. [Електронний ресурс] URL : <https://dokumen.pub/qdownload/python-data-science-handbook-essential-tools-for-working-with-data-2nbsped-1098121228-9781098121228.html>
3. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних / Я. В. Іванчук, В. І. Месюра, А. А. Яровий, О. Д. Манжілевський – Вінниця : ВНТУ, 2021. – 69 с. [Електронний ресурс] URL : <https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/36454/120997.pdf?sequence=2&isAllowed=y>

4. Методичні вказівки до лабораторної роботи "Основи роботи з бібліотекою Pandas" з курсу "Обробка даних Python" [Електронний ресурс] : для студентів спец. 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 124 Системний аналіз / уклад.: С. М. Коваленко, С. В. Коваленко ; Нац. техн. ун-т "Харків. політехн. ін-т". – Електрон. текст. дані. – Харків, 2022. – 44 с. – URI: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/61382>
5. Методичні вказівки до лабораторної роботи "Основи роботи з бібліотекою Matplotlib" з курсу "Обробка даних Python" [Електронний ресурс] : для студентів спец. 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 124 Системний аналіз, 126 Інформаційні системи і технології / уклад.: С. М. Коваленко, С. В. Коваленко, О. В. Шматко ; Нац. техн. ун-т "Харків. політехн. ін-т". – Електрон. текст. дані. – Харків, 2021. – 28 с. – URI: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/55469>
6. Методичні вказівки до лабораторної роботи "Основи роботи з бібліотекою NumPy" з курсу "Обробка даних Python" [Електронний ресурс] : для студентів спец. 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 124 Системний аналіз, 126 Інформаційні системи і технології / уклад.: С. М. Коваленко, С. В. Коваленко, О. В. Шматко ; Нац. техн. ун-т "Харків. політехн. ін-т". – Електрон. текст. дані. – Харків, 2021. – 48 с. – URI: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/55468>
7. Методичні вказівки до лабораторної роботи "Основи роботи в середовищі Jupyter Notebook" з курсу "Обробка даних Python" [Електронний ресурс] : для студентів спец. 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 124 Системний аналіз, 126 Інформаційні системи і технології / уклад.: С. М. Коваленко, С. В. Коваленко, О. В. Шматко ; Нац. техн. ун-т "Харків. політехн. ін-т". – Електрон. текст. дані. – Харків, 2021. – 28 с. – URI: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/55466>
8. Навчальний посібник для опанування студентами способів розв'язання задач з функціонального аналізу мовою Python. Частина 1 / Б. І. Мокін, В. Б. Мокін, О. Б. Мокін. – Вінниця : ВНТУ, 2022. – 124 с. [Електронний ресурс] URL : <https://iq.vntu.edu.ua/repository/getfile.php/5097.pdf>
9. Костюченко А. О. Основи програмування мовою Python: навчальний посібник. Ч.: ФОП Балакіна С. М., 2020. 180 с. [Електронний ресурс] URL : <https://epub.chnpu.edu.ua/jspui/bitstream/123456789/5584/1/%D0%9E%D1%81%D0%BD%D0%BE%D0%B2%D0%B8%20%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F%20%D0%BC%D0%BE%D0%B2%D0%BE%D1%8E%20Python.pdf>

Додаткова література

1. Lutz M. Learning Python 5th Edition. O'Reilly Media, Inc., 2013. 1600 p. [Електронний ресурс] URL : <https://edu.anarcho-copy.org/Programming%20Languages/Python/Learning%20Python.%205th%20Edition.pdf>
2. Lutz M. Python Pocket Reference. O'Reilly Media, Inc., 2014. 264 p. [Електронний ресурс] URL : <http://symbiod.com/pdf/Tech/Python%20Pocket%20Reference%204E.pdf?view=FitH>
3. Попередня обробка та аналіз даних: лабораторний практикум для студ. спеціальності 113 «Прикладна математика» / Уклад.: Н. Е. Кондрук. Ужгород: УжНУ, 2023. – 41 с. [Електронний ресурс] URL : <https://dspace.uzhnu.edu.ua/jspui/bitstream/lib/53297/1/%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D0%B8%D1%87%D0%BA%D0%B0%20%283%29.pdf>
4. NumPy. [Electronic resource]. Access mode: <https://numpy.org/>
5. The Jupyter Notebook. [Electronic resource]. Access mode: <https://jupyter.org/>
6. Pandas. [Electronic resource]. Access mode: <https://pandas.pydata.org/>
7. Matplotlib: Visualization with Python. [Electronic resource]. Access mode: <https://matplotlib.org/>
8. Seaborn. [Electronic resource]. Access mode: <https://seaborn.pydata.org/>
9. Scikit-learn Machine Learning in Python. [Electronic resource]. Access mode: <https://scikitlearn.org/stable/>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

Підсумкова оцінка з дисципліни розраховується як середня з кількох складових, що враховують оцінки кожного виду контролю:

- поточний контроль (2 контрольні роботи по 30 % кожна): 60 % семестрової оцінки;
 - лабораторні роботи: 20 % семестрової оцінки;
 - самостійна робота: 20 % семестрової оцінки;
- Якщо здобувач протягом семестру склав усі теми, то підсумкова оцінка може бути виставлена до початку сесії як результат накопичення оцінок. Здобувач за своїм бажанням може підвищити отриману оцінку, виконавши письмове завдання (два запитання з теорії + розв'язання двох практичних задач).

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

Погодження

Силабус погоджено

30.08.2024

Завідувач кафедри
Юрій ДОРОФЄЄВ

30.08.2024

Гарант ОП
Юрій ПАРЖИН